Jumping out of context – Jumping out of tone

Kalmanovitch, Yshai

Abstract: The study presented here investigates how different degrees of conceptual integration, as predicted by direct and indirect reported speech, effect speakers' prosodic behaviour in interaction. It makes a threeway distinction between authentic direct, false direct and indirect reported speech and predicts that greater conceptual distance, predicted by authentic direct reported speech, will lead to a greater deviation from the speaker's prosodic habitus with respect to register shifts, while the greater conceptual integration assumed in indirect reported speech will lead to greater prosodic integration.

Originally published at:
Kalmanovitch, Yshai (2005). Jumping out of context – Jumping out of tone. In: International Congress of Phonetic Sciences, Glasgow, 10 August 2005 - 14 August 2005, University of Glasgow.

JUMPING OUT OF CONTEXT – JUMPING OUT OF TONE

Yshai Kalmanovitch

Phonetics Laboratory, Department of Comparative linguistics, University of Zurich, Switzerland
yshai.kalmanovitch@uzh.ch

ABSTRACT

The study presented here investigates how different degrees of conceptual integration, as predicted by direct and indirect reported speech, effect speakers' prosodic behaviour in interaction. It makes a three-way distinction between authentic direct, false direct and indirect reported speech and predicts that greater conceptual distance, predicted by authentic direct reported speech, will lead to a greater deviation from the speaker's prosodic habitus with respect to register shifts, while the greater conceptual integration assumed in indirect reported speech will lead to greater prosodic integration.

Keywords: reported speech, conceptual alignment, prosodic shift, speech in conversation.

1. INTRODUCTION

In interaction, probably the most important assignment interlocutors have when communicating with each other is to coordinate and keep track of their common knowledge, or what Sperber and Wilson [19] refer to as the *mutual cognitive environment.* Thus, if Mary asks Paul to meet her "at the pub", she must either assume or otherwise make sure that Paul knows to which pub she refers.

While coordinating the mutual knowledge regarding a reference to a specific location is a relatively simple task, other tasks may prove to be more complicated. Thus, if Mary does not want to wait alone at the agreed location, she must make sure that Paul indeed sees it in his interest to meet her, and that this interest overcomes other interests he may have. In order to come to such an agreement, interlocutors must constantly negotiate their different assumptions, engaging themselves in a continuous process of conceptual alignment.

Different studies show that conceptual alignment in interaction leads also to alignment of the representational means and to increased coherence of the perceptual environment, which makes communication more efficient [15, 16]. This is further supported by many studies showing that prosodic discontinuity – or a deviation from the expected prosodic habitus – is perceived as reflecting also discontinuity in discourse [1, 5, 9, 20].

Prosodic discontinuity is often associated also with reported speech, which is used when reporting on speech acts in contexts other than the spatiotemporal context of the current interaction. Studies found tone and register reset, pauses, rhythm changes, increase and decrease of intensity and changes of voice quality to occur often in direct reported speech [2, 3, 5, 8, 9, 10, 11, 12, 13, 14]. Such prosodic shifts are often accounted for by the need to disambiguate reported speech structures [5, 12] or as a mean to comment on the reported speaker or the reported situation [3, 5, 8].

Most of those studies compared between direct and indirect reported speech, according to a traditional grammatical distinction. Thus, direct reported speech enacts the reported speech act as it would have been uttered by the reported speaker in the reported spatiotemporal context, as the reporting speaker takes the perspective of the reported speaker. On the other hand, indirect reported speech is uttered from the perspective of the reporting speaker in the current context in which it is reported, and requires a deictic shift (which become overt when the language morphology and syntax enable that).

Yet, the prosodic shifts associated with this distinction are considered as neither unique nor indexical to reported speech, and often correspond with the actual content of the reported speech act [12, 17]. In that respect, it is interesting to observe that direct and indirect reported speech differ also in their contribution to the process of conceptual alignment. This can be demonstrated in the best way observing the reporting of ceremonial speech acts. Thus, when reported by a current speaker in a current context, the two utterances *Tom told Jerry: "I sentence you to 10 years"* and *Tom told Jerry that he is sentencing him to 10 years*, reflect different degrees of commitment of the reporting speaker to the proposition expressed by the reported speech act, and thus require different degrees of modification of the mutual cognitive environment in order to reach conceptual alignment.

Thus, given a certain degree of tolerance, in direct reported speech the reporting speaker is perceived as committing only to the actual utterance of the reported words in the reported context, regardless of the validity and the effect of the reported speech in the reported context [7]. Without any clear commitment to the proposition expressed by the reported speech act, its validity in the current context

is very weak, thus hardly requires any modification of the mutual cognitive environment.

When using indirect reported speech on the other hand, the reporting speaker commits himself to the validity of the reported speech act and to its actual effect in the real world. This implies much greater effect in the process of conceptual alignment. In order for such potential effect to take place and not to be rejected by the hearer, the speaker needs to assume the smallest possible conflict between the negotiated assumption – the effect of the reported speech act in the real world – and other assumptions in the mutual cognitive environment. Thus, the proposition implied by an utterance in indirect reported speech must be well integrated in the interactional context.

Coming back to direct reported speech, however, there are frequent cases in which speakers use direct reported speech also when it is clear that the reported words were never uttered at all. Such cases may include reporting of generic communicational behaviour of a reported speaker or describing a certain state of mind of the reported speaker [17]. In other cases speakers may enact fictional dialogues [3, 17]. The degree of conceptual integration in those cases may vary. Thus, reporting generic behaviour may assume some information embedded in the mutual cognitive environment about the reported speaker, while reporting one's own thoughts will require only very little integration, if any at all. The enacting of hypothetic dialogues may call for cooperation of other interlocutors, which may themselves take part in the enacting as well, thus aligning to the reporting speaker [3, 17], reflecting very high degree of conceptual alignment.

In accordance with the hypotheses discussed earlier regarding the association between conceptual and prosodic integration in interaction, the current study investigates the hypothesis that higher conceptual integration will lead to higher prosodic integration and vice versa. The current study concentrates on register shifts in simple prototypical direct and indirect reported speech. Hybrid quotations [7] and enacted dialogues, which often lead to what can be described as voice polyphony [10] are thus excluded from the analysis.

In the current study reported speech in spoken Modern Israeli Hebrew will be investigated. The characteristics of reported speech in spoken Modern Israeli Hebrew are quite similar to those in English. It is presented with a quotative frame preceding the reported material, which contains a quotative verb or a quotative particle and often an overt marking of the reported speaker – and to lesser extent also of the addressee of the reported speech act. Indirect reported speech presents deictic shift and typically begins with a clitic /ʃe/ (similar to *that* in English) [13, 21].

A three-way distinction between authentic direct reported speech, false direct reported speech (which refers to reporting of speech that was never uttered, as discussed earlier) and indirect reported speech will allow the investigation of the prosodic behaviour in the three categories as a function of the expected contribution and effect of each category in the process of conceptual alignment. In accordance with the hypothesis investigated in this study, it is predicted that authentic direct reported speech, which is assumed to require the least integration in the process of conceptual alignment, will show the greatest shift from prosodic habitus. In contrast, indirect reported speech, which is assumed to be fully integrated in the mutual cognitive environment, will show the smallest shift, if at all. False reported speech, showing different intermediate degrees of conceptual integration is accordingly expected to show intermediate and possibly mixed tendencies.

2. DATA AND ANALYSIS

The data were collected from a corpus of natural conversational speech, designed for the investigation of speech accommodation in interaction. It contained 4 dyadic interactions between a single female speaker and 4 other individual speakers, two males and two females, all at mid. Twenties. All speakers were native speakers of spoken Modern Israeli Hebrew.

143 utterances including reported speech were extracted, of those 51 included authentic direct reported speech (aDRP), 52 false direct reported speech (fDRP) and 40 indirect reported speech (IRP). Since discontinuity in the flow of information may serve as an independent factor evoking register reset [6, 9], narrative segments immediately preceding and following the reported speech segment were included only if they constituted a coherent informational unit together with the reported material. Thus, while the narrative, which precedes the reported speech and by definition included at least the quotative frame (see below), seems obligatory, only 78 of the utterances included also a narrative segment immediately following it (31 aDRP, 32 fDRP and 15 IRP).

Using PRAAT [4], the peaks and lows in the extracted pitch contour of each phrase were labelled and manually corrected by the author, which is a trained phonetician and a native speaker of spoken Modern Israeli Hebrew. To avoid micro-intonation influence, target tones were measured only at the vowel of the syllable carrying the target tone, at the intensity peak of the estimated f0-band (75-500 Hz)

of the labelled vowel, which was automatically calculated using a script.

As mentioned above, the quotative frame was included in the narrative phrase immediately preceding the reported speech. The quotative frame is often described in the literature as part of the same intonation unit as the reported speech. However, at least in direct reported speech evidence seems to speak against such an account [10]. Major pauses, which are often taken as an indication of strong boundaries [1, 20], setting off the quoted material from the quotative frame, are also reported in the literature [13]. They were also observed in the current study, in which pauses longer than 100ms immediately preceding the reported material were found in 56% of the utterances containing direct reported speech (only a single case (2.5%) was detected in IRS).

Consequently, the boundaries between the reported material and the narrative segments were set at the onset and offset of the reported material itself, with the immediately preceding narrative phrase including the quotative frame, and often being constituted by it alone. The boundary in indirect reported speech was set immediately before the clitic /ʃe/, which often merged with the first syllable of the reported material, especially if this started with a vowel.

3. RESULTS

6 utterances were excluded due to background noise which did not allow reliable measuring (1 fDRS, 5 IRS). The differences of the peaks and the lows between the reported speech segments and the narrative segments immediately preceding and following them were calculated by subtracting the values of the target tones in the narrative segment from those in the reported speech segment (measured in semi tones) for each individual utterance. Thus a positive value means higher peaks and/or lows in the reported speech segments, while negative values reflect higher peaks and/or lows in the narrative segments.

Table 1 shows the mean differences of the peaks and the lows between reported speech segments (RS) and narrative segments immediately preceding them (N1), with the respective paired-samples t-test (two-tailed).

A one-way ANOVA using Brown-Forsythe test to overcome the imbalanced data found significant main effect for reported speech category for peaks ($F_{BF}(2,*133)=4.978$, p=.008) as well as lows ($F_{BF}(2,*107)=4.55$, p=0.013). Levene test confirmed the homogeneity of variance. Planned contrasts found significant mean differences between aDRS+fDRS and IRS for peaks (t(134)=3.078, p=.003) and lows (t(134)=2.831, p=.005), but not between aDRS and fDRS (for peaks and lows respectively: t(134)=0.274, n.s. and t(134)=1.249, n.s.).

Table 1: comparing registers of the reported material (RS) and the preceding narrative (N1)

	Peaks: RS-N2	Lows: RS-N2
aDRS	1.929 (±4.447) T(50)=3.098 p=.003	0.244 (±3.268) T(50)=0.753, n.s.
fDRS	1.7 (±4.351) T(50)=2.791 p=.007	-0.512 (±3.303) T(50)=-1.108 n.s.
IRS	-0.729 (±3.635) T(34)=-1.186 n.s.	-1.964 (±3.88) T(34)=-2.994 P=.005

Thus, aDRS shows a slightly higher register than the narrative immediately preceding it. Taking in account that the quotative frame by definition is expected to end with a continuous boundary tone, which means in Hebrew that the hypothetical lowest target tone will be avoided [18], and that the reported speech segments is very likely to end with a terminal boundary tone [15], the raised register in aDRS may be perceived as higher than initially seen from the statistics. This fact may also account for the lower register observed for IRS, with the significantly lower lowest target tones.

A similar analysis was carried out comparing reported speech (RS) and the narrative segments immediately following it (N2). Table 2 summarizes the results (note that the probability p=.04, found for the lows differences in IRS, is only marginally significant).

Table 2: comparing register of the reported material (RS) and the following narrative (N2)

	Peaks: RS – N1	Lows: RS – N1
aDRS	2.996(±3.203) T(30)=5.207 P<.001	1.455(±3.264) T(30)=2.482 P=.019
fDRS	1.062(±3.567) T(31)=1.683 n.s.	0.674(±2.667) T(31)=1.374 n.s.
IRS	1.433(±4.297) T(14)=1.292 n.s.	1.944(±3.303) T(14)=2.267 P=.04

A one-way ANOVA found no effect of category for neither mean peak differences ($F_{BF}(2,*43)=2.203$, n.s.) nor for mean lows differences ($F_{BF}(2,*52)=0.954$, p=.059).

Combining both comparisons, IRS can be described as showing a steady decline of register in the sequence N1-RS-N2. In comparison, aDRS shows a noticeable register-shift upwards in the reported speech segment relative to the narrative segments, while fDRS shows intermediate tendencies.

To make sure this picture was not biased by the lack of final narrative segment in many of the utterances considered in table 1, this analysis was repeated once more, this time only for the utterances that were considered also in table 2. The results are summarized in table 3.

Table 3: comparing register of the reported material and the preceding narrative (N1) including only the utterances with following narrative.

	Peaks: RS-N1	Lows: RS –N1
aDRS	3.146 (±4.057) T(30)=4.646 P<.001	0.729(±3.518) T(30)=1.146 n.s.
fDRS	0.509 (±3.785) T(31)=0.76 n.s.	0.099(±2.941) T(31)=0.224 n.s.
IRS	0.25 (±4.294) T(14)=0.226 n.s.	-1.098(±4.398) T(14)=0.967 n.s.

A one-way ANOVA found a main effect for category for mean peak differences (F_{BF} (2,*49)=4.351, p=.018), but not for mean lows differences (F_{BF} (2,*36)=1.256, n.s.). Levene test confirmed the homogeneity of variance. Unlike the analysis of the data in table 1, the Planned contrasts found significant difference between mean peak differences in aDRS and fDRS (T(75)=-2.697, p=.009) but not between both categories of direct reported speech and the category of indirect reported speech. Nevertheless, the general trends in table 1 and table 3 are very similar.

4. DISCUSSION

The results described above support the main hypothesis investigated in this study. Indirect reported speech, which expresses greater commitment of the speaker to the effect of the reported speech act on the process of conceptual alignment in the current interaction, is well integrated in a general declination line, which follows what can be described as the default prosodic behaviour within a coherent informational unit [9].

On the other side, authentic direct reported speech shows a noticeable shift from this expected prosodic behaviour, as the reported material is uttered in a higher register than the narrative segment immediately preceding it. Thus, a prosodic discontinuity is expected to be perceived by the listener, which is often taken as a sign for informational discontinuity [6, 9]. The perception of discontinuity is expected to be increased also by the frequent presence of a long pause immediately preceding the onset of the reported material, as discussed above.

False direct reported speech, despite its grammatical similarities with authentic reported speech, shows what seems to be an intermediate trend between the two categories described above. Thus, the register shift described in authentic direct reported speech is unlikely to be described as inherent to the general distinctive grammatical structure of direct reported speech.

The results also seem to replicate findings in earlier studies, referring to a wider range in direct reported speech [10], as the peaks seem to show greater shift upwards than the lows. This observation may be the consequence of the relatively frequent use of vocatives and emphatic particles (e.g. *oh my god*, *wow*, etc.), which are completely absent from indirect reported speech and often seem to call for a high jump upwards. Thus, the wider range may reflect not necessarily a typical prosodic behaviour unique to direct reported speech, but rather be associated with some unique lexical characteristics associated with it.

It should also be pointed out that in both categories of direct reported speech the narrative immediately following the reported material, though shifting downwards, doesn't seem to come back and continue the hypothetic declination of the baseline of the initial narrative (otherwise a much greater shift downwards would have been expected in comparison with the shift observed in indirect reported speech). Thus, the observed register shift in authentic – and to some extent in false – direct reported speech may not be really a register shift, but rather a register reset. With the lack of conclusive statistical evidence this question needs further investigation.

5. REFERENCES

[1] Amir, N., Silber-Varod, V., Izre'el, S. 2004. Characteristics of intonation unit boundaries in spontaneous spoken Hebrew-perception and acoustic correlates. In *Speech Prosody 2004, International Conference*. 677-680.

[2] Bertrand, R., Espesser, R. 2002. Voice Diversity in Conversation: a Case Study. In *Speech Prosody 2002, International Conference*. 171-174

[3] Bolden, G. 2004. The quote and beyond: defining boundaries of reported speech in conversational Russian. *Journal of pragmatics*, *36*(6), 1071-1118.

[4] Boersma, P., Weenink, D. 2014. *PRAAT: Doing Phonetics by Computer* (Version 5.3.78). www.praat.org

[5] Couper-Kuhlen, E. 1999. Coherent voicing: On prosody in conversational reported speech. *Pragmatics & beyond. New series*, *63*, 11-33.

[6] Couper-Kuhlen, E. 2001. Intonation and discourse: Current views from within. In: Shiffrin, D., Tannen, D., Hamilton, H (eds.). *The handbook of discourse analysis*. Oxford: Blackwell. 13-34.

[7] De Brabanter, P. 2010. The semantics and pragmatics of hybrid quotations. *Language and Linguistics Compass*, *4*(2), 107-120.

[8] Günthner, S. 1999. Polyphony and the 'layering of voices' in reported dialogues: An analysis of the use of prosodic devices in everyday reported speech. *Journal of pragmatics*, *31*(5), 685-708.
[9] Hirschberg, J., Grosz, B. 1992. Intonational features of local and global discourse structure. In *Proceedings of the workshop on Speech and Natural Language* (pp. 441-446). Association for Computational Linguistics
[10] Jansen, W., Gregory, M. L., Brenier, J. M. 2001. Prosodic correlates of directly reported speech: Evidence from conversational speech. In *ISCA Tutorial and Research Workshop (ITRW) on Prosody in Speech Recognition and Understanding*.
[11] Kasimir, E. 2008. Prosodic correlates of subclausal quotation marks. *ZAS Papers in Linguistics*, 49, 67-78
[12] Klewitz, G., Couper-Kuhlen, E. 1999. Quote-unquote: the role of prosody in the contextualization of reported speech sequences. *Pragmatics*, 9(4), 459-485.
[13] Malibert, I.I., Vanhove, M. 2015. Quotative constructions and prosody in some Afroasiatic languages: Towards a typology. In: Mettouchi, A., Vanhove, M., Caubet. D. (eds), *Corpus-based Studies of Lesser-described Languages: The CorpAfroAs corpus of spoken AfroAsiatic languages.* Amsterdam: Benjamins, 117-169.
[14] Oliveira Jr, M., Cunha, D. A. 2004. Prosody as marker of direct reported speech boundary. In *Speech Prosody 2004, International Conference*.
[15] Pickering, M. J., Garrod, S. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, *27*(02), 169-190.
[16] Pickering, M. J., Garrod, S. 2013. An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(04), 329-347.
[17] Sams, J. 2010. Quoting the unspoken: An analysis of quotations in spoken discourse. *Journal of Pragmatics*, *42*(11), 3147-3160.
[18] Silber-Varod, V., Kessous, L. 2008. Prosodic boundary patterns in Hebrew: A case study of continuous intonation units in weather forecast. In *Proceedings of the Speech Prosody 2008 Conference*, 265-268.
[19] Sperber, D., Wilson, D. 1986. *Relevance: Communication and Cognition*. Oxford: Blackwell.
[20] Swerts, M. 1997. Prosodic features at discourse boundaries of different strength. *The Journal of the Acoustical Society of America*, *101*(1), 514-521.
[21] Zuckermann, G. 2006. Direct and indirect speech in straight-talking Israeli. *Acta Linguistica Hungarica*, *53*(4), 467-481.